

### Das Programmsystem LVPLS für Pfadmodelle mit latenten Variablen

Lohmöller, Jan-Bernd

Veröffentlichungsversion / Published Version  
Zeitschriftenartikel / journal article

Zur Verfügung gestellt in Kooperation mit / provided in cooperation with:  
GESIS - Leibniz-Institut für Sozialwissenschaften

#### Empfohlene Zitierung / Suggested Citation:

Lohmöller, J.-B. (1984). Das Programmsystem LVPLS für Pfadmodelle mit latenten Variablen. *ZA-Information / Zentralarchiv für Empirische Sozialforschung*, 14, 44-51. <https://nbn-resolving.org/urn:nbn:de:0168-ssoar-206293>

#### Nutzungsbedingungen:

Dieser Text wird unter einer Deposit-Lizenz (Keine Weiterverbreitung - keine Bearbeitung) zur Verfügung gestellt. Gewährt wird ein nicht exklusives, nicht übertragbares, persönliches und beschränktes Recht auf Nutzung dieses Dokuments. Dieses Dokument ist ausschließlich für den persönlichen, nicht-kommerziellen Gebrauch bestimmt. Auf sämtlichen Kopien dieses Dokuments müssen alle Urheberrechtshinweise und sonstigen Hinweise auf gesetzlichen Schutz beibehalten werden. Sie dürfen dieses Dokument nicht in irgendeiner Weise abändern, noch dürfen Sie dieses Dokument für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, aufführen, vertreiben oder anderweitig nutzen.

Mit der Verwendung dieses Dokuments erkennen Sie die Nutzungsbedingungen an.

#### Terms of use:

This document is made available under Deposit Licence (No Redistribution - no modifications). We grant a non-exclusive, non-transferable, individual and limited right to using this document. This document is solely intended for your personal, non-commercial use. All of the copies of this documents must retain all copyright information and other information regarding legal protection. You are not allowed to alter this document in any way, to copy it for public or commercial purposes, to exhibit the document in public, to perform, distribute or otherwise use the document in public.

By using this particular document, you accept the above-stated conditions of use.



## DAS PROGRAMMSYSTEM LVPLS FÜR PFADMODELLE MIT LATENTEN VARIABLEN

Jan-Bernd Lohmöller

Das diesjährige Frühjahrsseminar des Zentralarchivs war dem Studium zweier teils konkurrierender, teils komplementärer Ansätze der Pfadanalyse mit latenten Variablen gewidmet, dem LISREL- und dem PLS-Ansatz. In diesem Artikel soll — nach einer kurzen Gegenüberstellung der Methoden — das Programmsystem LVPLS (Latent Variables Path Analysis with Partial Least Squares Estimation, Lohmöller 1981) vorgestellt werden. Das Programmsystem LVPLS kann vom Zentralarchiv bezogen werden. In der nächsten Nummer der ZA-Information wird auf die konzeptionellen Unterschiede des PLS- und LISREL-Ansatzes näher eingegangen, und es wird an einem Beispiel die Ähnlichkeit und Unterschiedlichkeit der Ergebnisse gezeigt. (Ein Vorabdruck kann vom ZA angefordert werden.)

### PLS und LISREL

Die Pfadanalyse mit latenten Variablen hat im letzten Jahrzehnt eine weite Verbreitung gefunden, was nicht zuletzt an der Verfügbarkeit des LISREL-Programms von Jöreskog und Sörbom (1981) liegt. Karl Jöreskog hat drei Dinge entwickelt, die alle drei LISREL (Linear structural relations) heißen. Diese drei lassen sich zum Teil durch teilweise nützlichere Dinge ersetzen.

Das LISREL-Modell ist eine Kombination von Pfadmodell und Faktorenmodell. Eine Menge von latenten Variablen (LVn) wird als ursächlich im Sinne gemeinsamer Faktoren für eine Menge von manifesten Variablen (MVn) angenommen. Die Beziehungen der LVn untereinander sind als ein Pfadmodell gefaßt. Kombinationen von Pfad- und Faktorenmodell wurden etwa gleichzeitig (1973) von Jöreskog, von Keesling und von Wiley vorgestellt, wie Bentler (1980) in einem Überblicksartikel darlegt.

Soweit es um die Darstellung von Beziehungen zwischen den Modellvariablen geht, unterscheidet sich das LISREL-Modell nicht von den Entwicklungen, die Herman Wold unter den Namen "Soft modelling" oder "PLS modelling" vorgelegt hat (1975, 1982). Der Name "Pfadmodell mit latenten Variablen" (latent variables path model, LVPM) kann als neutrale und umfassende Bezeichnung verstanden werden. Die Unterschiede zwischen der LISREL- und PLS-Version des LV-Pfadmodells beginnen mit den Verteilungsannahmen. Im (klassischen) LISREL-Modell werden die Variablen als multivariat normal-

verteilt angenommen. Im PLS-Ansatz werden nur die bedingten Erwartungen (die Regressionslinien) spezifiziert; es werden also entschieden schwächere Annahmen gemacht.

Die LISREL-Methode der Parameterschätzung ist eine Maximum-Likelihood-Methode, und die LISREL-Schätzungen erfreuen sich damit der wünschenswerten Eigenschaft der Präzision von ML-Schätzungen. Der Aufwand an Rechenzeit ist allerdings für größere Modelle unerfreulich groß. Die PLS-Methode (Partial Least Squares) dagegen ist eine Kleinstquadratmethode, und der PLS-Algorithmus, der nichts als Regressions- und Korrelationsberechnungen enthält, ist relativ schnell.

Ein wesentlicher Unterschied zwischen PLS und LISREL betrifft die Zusatzannahmen, die das Modell schätzbar machen. In PLS werden die latenten Variablen als gewichtete Aggregate geschätzt; das impliziert, daß die LV-Werte (Faktorwerte) zum konstitutiven Modellbestandteil werden. PLS ist somit eine Generalisierung der Hauptkomponentenanalyse, während LISREL eine Generalisierung der Faktorenanalyse ist. Es gibt in LISREL keine LV-Werte, die mit den übrigen Modellparametern eine kohärente Lösung bilden. Mit den LV-Werten kann in PLS die gesamte Datenmatrix reproduziert (prädiziert) werden, während in LISREL die Kovarianzmatrix reproduziert wird. Darstellungen der PLS-Methode finden sich außer bei Wold und Lohmöller (alle Literaturangaben) auch bei Knepel (1980) und Fornell (1982); Gegenüberstellungen von PLS und LISREL bei Fornell und Bookstein (1982) und Jöreskog und Wold (1982b).

Das LISREL-Programm ist speicherplatz- und rechenzeitintensiv. Nicht immer konvergiert der Schätzalgorithmus, nicht immer sind die Schätzungen plausibel, nicht immer zeigen die Ergebnisse, was am Modell falsch ist. Das macht die Datenexploration und Modellentwicklung mit LISREL schwierig und teuer. Die Programme LVPLSX (Rohdatenmatrix als Eingabe) und LVPLSC (Kovarianzmatrix als Eingabe), Teile des Programmsystems LVPLS, benötigen weniger Computerressourcen.

#### Einige Argumente für LVPLS

Tafel 1 zeigt einen Betriebsmittelvergleich für mehrere kleine und große Modelle, die mit LISREL IV und LVPLS geschätzt wurden. Als erstes

Tafel 1 Betriebsmittelbedarf für LVPLSC und LISREL IV  
für verschiedene kleine und große Modelle

Modell	MV	LV	Par	KSB	CPU
Noonan & Wold (1977)	7	3	9 17	268 640	1 2
Augsburger Längsschnitt kleines Modell	12	4	14 29	594 1408	2 6
Holzinger & Swineford (LISREL IV Manual)	9	3	27 33	407 1280	4 13
Georg Lohmöller: Herz-Kreislauf-Modell	16	7	32 57	1010 3840	5 48*
Marjoribanks (s. Lohmöller 1979)	18	6	38 51 51	1310 3456 3712	5 177* 27
Lord's speed factor (Psychometrika 43:454)	18	6	42 59	1394 3840	6 86
Augsburger Längsschnitt, großes Modell	83	21	260 332	20858 87168	700 1800*
Experiment Alkoholwirkungswahrnehmung	94	4	106 .	13906 51456	9 sing*
Wingard	96	4	265 .	33520 .	(300)*
Schulte (1981)	96	26	287 .	26739 116736	360 sing*
Bartl, Unverdorben & Lohmöller Soziale Probleme im Grundwehrdienst	102	23	230 321	26266 96896	198 1800*
Gattringer (1980)	57	10	75	.	.
Engfer & Schneewind	82	13	103	147217	16

Die erste Zeile zu jedem Modell bezieht sich auf LVPLSC, die zweite Zeile auf das LISREL-Modell.

MV Anzahl der manifesten Variablen

LV Anzahl der latenten Variablen

Par Anzahl der Parameter des Modells

KSB Kernspeicherbedarf in Worten (=Kbyte x 128)

CPU Kernrechenzeit in Sekunden, Burroughs B7800

\* keine Konvergenz in 250 Iterationen oder Rechenzeitschranke

sing MV-Kovarianzmatrix ist singulär, keine LISREL-Schätzung möglich  
nicht bekannt



Beispiel sei das Modell über "Soziale Probleme im Grundwehrdienst" (Bartl, Unverdorben & Lohmöller 1981) herausgegriffen, das 102 MVn und 23 LVn umfaßt. Die erste Version des Modells mit unrestringierter Pfadmatrix und arbiträren Startwerten benötigte 15 Minuten Kernrechenzeit. Die endgültige Modellversion mit 230 Parametern in Pfad- und Gewichtsmatrix und mit vernünftigen Startwerten benötigte 26'266 Kernspeicherworte (=154 KByte), 41 Iterationen und 198 Sekunden Kernrechenzeit auf der Burroughs B7800. Das von LVPLSC generierte LISREL-Modell umfaßte 321 Parameter in Pfad-, Ladungs- und Residualkovarianzmatrizen. LISREL vollführte 4 steepest-descend-Iterationen in 30 Minuten; die Lösung war weit von den Konvergenz entfernt.

Als zweites Beispiel vergleiche man zwei konfirmatorische Faktorenanalysen, das "Experiment Alkoholwirksamkeit" (unveröffentlicht) mit 94MVn, das mit PLS in 9 Sekunden konvergierte, und den Versuch von Wingard (pers. Mitteilung;), eine Faktorenanalyse von 96 MVn mit LISREL zu rechnen: Nach 5 Minuten Kernrechenzeit auf einer IBM 3022 waren zwei steepest-descend-Iterationen vollbracht.

Auch wenn die wenigsten unserer Leser für Computerzeit zu zahlen haben: Es macht einen Unterschied, ob man sein Resultat nach 5 Minuten oder am nächsten Tag erhält. Neben dem simplen Argument der Rechenzeit seien einige Anwendungsbereiche aufgeführt, in denen die konzeptionellen Unterschiede eher für den PLS- als für den LISREL-Ansatz sprechen.

Kleine Stichproben. In kleinen Stichproben ( $N < 200$ ) wirken sich die (theoretischen) asymptotischen Optimalitätseigenschaften von ML-Schätzern noch nicht aus, und praktisch kann die Stichprobenkovarianzmatrix zu unpräzise sein, um mit dem LISREL-Programm zu sinnvollen Schätzungen zu kommen. Wold (1980, unveröffentlicht) zeigte, daß ein PLS-Modell für einen Datensatz mit  $N = 10$  Fällen und 27 Variablen prädiktive Kraft hatte.

Kleine finite Populationen. Untersuchungen über internationale Politik haben als Untersuchungseinheiten 150 Nationen, mehr gibt es nicht. Die Untersuchungseinheiten sind nicht unabhängig voneinander, und Verteilungsannahmen machen keinen Sinn. Hier sind alle Voraussetzungen der (klassischen) LISREL-Modellierung verletzt, aber keine der PLS-Modellierung.

Große Variablenmengen. Bei einem LISREL-Modell mit 50 manifesten Variablen wird die Grenze der Leistungsfähigkeit heutiger Computer



berührt. Wie Tafel 1 zeigt, liegen die Rechenzeiten für PLS-Modelle mit ca. 100 manifesten Variablen bei wenigen Minuten. Das bisher größte PLS-Modell hat an die 500 Variablen.

Exploration. Weil PLS nicht verlangt, daß das spezifizierte Modell "das wahre" ist, und weil wenig Computerzeit benötigt wird, ist PLS für die Exploration von Modellen weit besser geeignet als LISREL. Dieses ist besonders in solchen Anwendungsbereichen von Nutzen, in denen wenig substantielle Vorinformation vorhanden ist und in den Modellentwurf inkorporiert werden kann, beispielsweise in interdisziplinären Projekten.

Prädiktion, Prognose. Zielt eine Untersuchung darauf ab, für jeden einzelnen Fall in der Datenmatrix zu einer Entscheidung, z.B. über Schulzulassung, Einstellung, Therapiezuweisung, zukommen, dann muß die Prädiktion/Prognose des Schul-/Arbeits-/Therapieerfolges optimal sein. Hier ist ein PLS-Modell angezeigt.

#### Das Programmsystem LVPLS

LVPLS ist ein Programmsystem, das aus mehreren "Haupt"-Programmen besteht, die im Prinzip als lauffähige, selbständige Programme herausgetrennt werden können, die in LVPLS jedoch von einem Monitorprogramm aufgerufen werden.

LVPLSX erwartet als Eingabe eine Rohdatenmatrix. Die Datenmatrix, die im Kernspeicher abgelegt wird, muß frei von Missing Data sein. Das zu schätzende Modell muß dem Standard der grundlegenden Methode (Basic Design von Wold) genügen: Das innere Modell (Pfadmodell) ist ein rekursives (schleifenfreies) Pfadmodell, und im äußeren Modell ist jede LV an einen Block (=eine Menge) von MVn geknüpft. Die Gesamtmenge der eingelesenen Variablen zerfällt damit in disjunkte Blöcke, womit nicht ausgeschlossen ist, daß eine Variable zweimal eingelesen wird und in zwei verschiedenen Blöcken vorkommt. Das Programm druckt außer Pfad-, Ladungs-, Gewichts- und LV-Korrelationsmatrix sowie zusammenfassende Statistiken auch die LV-Werte, d.h. die Scores der Fälle auf den geschätzten LVn. Auf Anforderung wird für einen Block das Blindfolding durchgeführt, das in einem verteilungsfreien Kennwert für die prädiktive Relevanz des Modells und in Standarabweichungen für die Parameterschätzungen dieses Blocks resultiert. Dann berechnet das



Programm die Residualdatenmatrix, und es kann eine weitere Dimension der LVn und eine weitere Dimension der Parameter des gesamten Modells geschätzt werden. Am Schluß kann eine Kovarianzmatrix aller Residual- und aller latenten Variablen ausgedruckt werden.

LVPLSC erwartet als Eingabe eine Kovarianz- oder Korrelations- oder Rohprodukt- oder Kontingenzmatrix. Durch einen vorausgehenden Aufruf des Hauptprogramms LVPSCA kann eine mit Mittelwerten und Standardabweichungen eingegebene Korrelationsmatrix zurücktransformiert werden in eine Kovarianz- oder Rohproduktmatrix, und zwar getrennt für verschiedene Blöcke von MVn. Das zu schätzende Modell muß dem Standard der erweiterten Methode (Extended Design, Lohmöller 1979) genügen: Das innere Modell ist wiederum ein schleifenfreies Pfadmodell, aber im äußeren Modell können mehrere LVn pro Block (anders gesagt, mehrere Dimensionen einer LV), und von Block zu Block unterschiedliche Anzahlen von LVn, gleichzeitig geschätzt werden. Zum Zwecke der Modellüberprüfung druckt das Programm die Kovarianz der inneren Residualvariablen,  $Q = \text{cov}(u)$ , der äußeren Residualvariablen,  $C = \text{cov}(e)$ , die im Kovarianzstrukturmodell in der Regel a priori als diagonal angenommen werden, sowie die Kovarianzen der LVn und Residualvariablen,  $\text{cov}(Y;e)$  und  $\text{cov}(u;e)$ , die in LISREL a priori als Null angenommen werden. In einer Zusammenfassungsseite druckt LVPLSC eine Reihe von Statistiken, die eine Abschätzung der Auswirkungen von Modellmodifikationen und einen Vergleich mit anderen Schätzmethoden erlauben.

Auf Anforderung generiert LVPLSC ein vollständiges Eingabedeck für LISREL (IV oder V oder VI), das die PLS-Schätzungen als Startwerte enthält. Der Benutzer hat die Wahl zwischen drei Regeln, nach denen das PLS Modell in ein LISREL-Modell übersetzt wird, die NO-X-Regel, die FIXED-X-Regel, und eine dritte Regel, die auf ein Theorem von Dupacová und Wold (in Jöreskog & Wold 1982a) über die Identifizierbarkeit von ML-Modellen zurückgeht. Die beste Strategie zur Konstruktion von LISREL-Modellen besteht darin, die Standards der grundlegenden PLS-Methode einzuhalten: Schleifenfreie Pfadmodelle mit disjunkten Indikatoren sind immer identifiziert, wenn die Skalenfixierungen zugefügt werden. Der Benutzer, der PLS- und LISREL-Schätzungen vergleichen möchte, braucht nichts von der Bedienung des LISREL-Programms zu verstehen, wenn er LVPLSC als Präprozessor benutzt.

LVPLS3 erwartet als Eingabe die Kovarianzmatrix eines Datenquaders  $[x_{j_{in}}]$ , der aufgeblättert wurde zu einer Datentafel  $[x(jt)n]$ , worin



der Index  $n$  beispielsweise für eine Person steht, deren Attribut  $j$  zu einem Zeitpunkt  $t$  den Wert  $x_{jtn}$  hat. Das Modell der trimodalen Pfadanalyse mit latenten Variablen, das das trimodale Faktorenmodell und das LV-Pfadmodell als Spezialfälle abdeckt, wurde von Lohmöller und Wold (1980, 1982) vorgestellt.

LVPGEN generiert nach einem vorgegebenen LV-Pfadmodell eine Rohdaten-, Kovarianz- oder Korrelationsmatrix, und zwar für eine Population oder für eine Stichprobe. Die erzeugte Matrix kann im gleichen LVPLS-Lauf durch eines der drei Analyseprogramme LVPLSX, LVPLSC, LVPLS3 ausgewertet werden.

Programmdokumentation • Für die zugrundeliegenden Methoden und Algorithmen wird auf Wold (1982) und Lohmöller (1984) verwiesen. Die programmspezifische Information umfaßt:

- LVPLS 1.6 program manual (ca. 100 Seiten)
- LVPLS 1.6 instant manual (16 Seiten, auf Magnetband)
- LVPLS 1.6 implementation manual (30 Seiten, auf Magnetband)

Das Programmanual, derzeit direkt vom Zentralarchiv erhältlich, wird demnächst im Physica-Verlag Würzburg erscheinen.

Programmimplementation. LVPLS ist in Fortran geschrieben. In seinen wesentlichen Teilen genügt es einem sehr strengen Fortran-Standard und kann mit fast allen Fortran-IV- und Fortran-77-Compilern übersetzt werden. LVPLS läuft auf Maschinen der Typen Burroughs, CDC, IBM, DEC 10, DEC 20, TR 440, Univac, Siemens BS2000 und BS3000 (Fujitsu), Vax, Prime. Aus etwa 30 Universitätsrechenzentren liegen Rückmeldungen über erfolgreiche Implementationen vor.

#### Literatur

- Bartl, P., Unverdorben, K., Lohmöller, J.B. (1981) Soziale Probleme im Grundwehrdienst: Eine Pfadanalyse zu Alkohol- und Suizidproblemen (Forschungsbericht 81.01 Fachbereich Pädagogik). München: Hochschule der Bundeswehr.
- Bentler, P.M. (1980) Multivariate analysis with latent variables: Causal modeling. Annual Review of Psychology, 31: 419-456.
- Engfer, A. (1982) Bedingungen und Auswirkungen harten elterlichen Strafens (Dissertation). Trier: Universität.
- Fornell, C. (Ed.) (1982) A second generation of multivariate analysis, vol. 1 - measurement and evaluation. New York: Praeger.
- Fornell, C. & Bookstein, F.L. (1982) Two structural equation models: LISREL and PLS applied to consumer exit-voice theory. Journal of Marketing Research, 19: 440-452.
- Gattringer, H. (1980) Die Entwicklung des Leistungsmotivs und dessen Bedeutung für die Schulleistungsprognose (Dissertation). Universität Salzburg: Philosophische Fakultät.





- Jöreskog, K.G. & Sörbom, D. (1981) LISREL V - Analysis of linear structural relationships by maximum likelihood and least Squares methods (Research Report 81-8). Uppsala: University Department of Statistics.
- Jöreskog, K.G. & Wold, H. (Eds.) (1982a) Systems under indirect Observation: Causality, structure, prediction (2 vols.). Amsterdam: North Holland.
- Jöreskog, K.G. & Wold, H. (1982b) The ML and PLS techniques for modeling with latent variables: Historical and comparative aspects. In K.G. Jöreskog & H. Wold (Eds.), Systems under indirect Observation: Causality, structure, prediction (2 vols.). Amsterdam: North Holland (vol. 1, p.263-270).
- Knepel, H. (1980) Modelle mit unbeobachteten Variablen: der PLS-Ansatz. Statistische Hefte, 2<sup>4</sup>:248-279.
- Lohmöller, J.B. (1979) Estimating parameters of **linear** structural relation models under partial least-squares criteria (Forschungsbericht 79.01 Fachbereich Pädagogik, Hochschule der Bundeswehr München). Microfiche ERIC ED 178 600.
- Lohmöller, J.B. (1981) LVPLS 1.6 program manual: Latent variables path analysis with partial least Squares estimation (Forschungsbericht 81.04 Fachbereich Pädagogik). München: Hochschule der Bundeswehr.
- Lohmöller, J.B. (1984) Path models with latent variables and Partial Least Squares (PLS) estimation (Dissertation, Hochschule der Bundeswehr München, Fachbereich Pädagogik, 1983). Würzburg: Physica-Verlag (in Vorbereitung).
- Lohmöller, J.B. & Wold, H. (1982) Pfad- und faktorenanalytische Ansätze zur differentiellen Entwicklungsbeschreibung: Die trimodale Pfadanalyse mit latenten Variablen (Band 1, S. 36-43). In R. Oerter (Hrsg.), Bericht über die 5. Tagung Entwicklungspsychologie in Augsburg, 21.-23. Sept. 1981 (2 Bände). Augsburg: Universität.
- Schneewind, K. A. , Beckmann, M., & Engfer, A. (1983) Eltern und Kinder: Umwelteinflüsse auf das familiäre Verhalten. Stuttgart: Kohlhammer.
- Wold, H. (1975) Path models with latent variables: The NIPALS approach. In H.M. Blalock & al (Eds.), Quantitative sociology: International perspectives on mathematical and Statistical modeling. New York: Academic (p.307-357).
- Wold, H. (1982) Soft modeling: The basic design and some extensions. In K.G. Jöreskog & H. Wold (Eds.), Systems under indirect Observation: Causality, structure, prediction (2 vols.). Amsterdam: North Holland (vol.2, p.1-54).

Jan-Bernd Lohmöller  
Zentralinstitut für sozialwissenschaftliche Forschung  
Arbeitsbereich Vergleichende Faschismusforschung  
Freie Universität Berlin  
Sarrazinstraße 11-15  
D-1000 Berlin 41